

ABSTRACT AFRICAGIS 2017

ESTIMATING MAIZE GRAIN YIELD IN SCARCE FIELD-DATA ENVIRONMENT: AN APPROACH COMBINING REMOTE SENSING AND CROP MODELLING IN BURKINA FASO

Leroux Louise, Christian Baron, Mathieu Castets, Maria-José Escorihuela, A. Aziz Diouf, Agnès Bégué, Danny Lo Seen

INTRODUCTION

With more than a billion tons per year, maize is the most produced crop in the world and is thus considered as key in support of global food security. In West Africa, maize is a staple crop and plays a central role in fulfilling the food requirements of the population (Chivasa et al., 2017) whose average consumption is estimated at more than 30 kg/capita/year in 2013 by FAOSTAT. However, in the region, important climate and demographic trends combine to worsen an already difficult situation. Population growth is expected to be higher than rates of yield improvement, which suggests a decline in per capita food production in the coming years (Ray et al., 2013). In that context, timely and reliable information on maize crop yields and its spatial variability is urgently needed in order to provide timely estimates of food shortage to support early warning food security systems.

There exist several methods for crop yield estimates from plot to continental scale. Traditional methods such as field-based surveys are theoretically the most appropriate way to obtain reliable maize yield estimates but they also have significant weaknesses, among which the cost, in terms of time and labor, the inaccessibility of certain areas, and the difficulty to upscale to large areas (Burke and Lobell, 2017). That is why in West Africa such data are generally lacking and/or is often unreliable, however they are necessary. Another way to estimate crop yield is to use crop growth models that incorporate ecophysiological processes. However, their use is limited by the availability of input data at the field scale over large areas. Due to close relationships between some vegetation indices and certain biophysical crop parameters, remote sensing has been successfully used for yield estimation of large homogeneous crop plots in developed countries and some promising results have been reached for African smallholder agriculture (Leroux et al., 2016). These approaches mainly rely on empirical relationships between remote sensing indices and agricultural national statistics, due to the scarcity of reliable in-situ yield measurements. Still, agricultural statistics are generally available 3 months after the end of the cropping season and thus don't allow to have timely yield estimations.

To overcome the lack of in situ yield and timely agricultural statistical data, we propose in this study to use a crop model to generate different components of yield as proxies to observed data - referred hereafter as an "uncalibrated approach" -, and to combine them with remote sensing data. Our specific objectives are to (1) build a remote sensing-based yield model relying on ecophysiological processes implemented in crop growth models, and (2) to compare the performance of linear and non-linear statistical models to estimate maize crop yield in an African smallholder agriculture context. The remote sensing-based yield model is based on the estimation of vegetative biomass at flowering stage, and on the estimation of a reducing factor of yield which is a water stress coefficient using statistical methods and vegetation indices, land surface temperature and soil surface moisture time series of low resolution images. This approach is applied to estimate maize grain yield in the south-west of Burkina Faso.

DATA AND PREPROCESSING

STUDY AREA

The study is conducted from 2011 to 2016 in the Koumbia village, located in the Tuy province in south-western Burkina-Faso. The climate is Sudanian, characterized by a uni-modal rainfall season, with annual rainfall ranging from 750 mm to 1100 mm and a rainy season lasting from June to September. As in most of West Africa, the region is characterized by a high inter-annual and intra-annual rainfall variability that can affect food security, with the population largely relying on rainfed crops. Maize and cotton are the main crops cultivated in the Koumbia village, covering about 90% of the cultivated area (Diarisso et al., 2015).

FIELD DATA

Field data was collected in 2014, 2015 and 2016 for 3 villages located in the study area. A network of 114 maize fields under farmer conditions was surveyed to monitor agricultural practices and vegetation parameters (e.g. aerial biomass, final grain yield). Field data were used to verify the robustness of the remote-sensing maize yield model against independent data.

REMOTE SENSING DATA

Three remote sensing products were used for the purpose of this study: (1) NDVI (Normalized Difference Vegetation Index) from the MODIS MOD13Q1 product (16-day, 250 m spatial resolution), (2) LST (Land Surface Temperature) from the MODIS MOD11A2 product (8-day, 1 km spatial resolution) and (3) SSM (Soil Surface Moisture) from SMOS (daily, ~40 km spatial resolution). A smoothing filter was applied to NDVI and LST MODIS time series using a Savitsky-Golay algorithm in order to improve the quality of the time series. Soil Moisture is highly variable spatially, thus we applied a disaggregation approach in order to obtain more relevant soil moisture information for the monitoring of rainfed crops in heterogeneous agricultural landscapes. The SMOS SSM data were disaggregated at a 1-km spatial resolution using the DISPATCH method (Merlin et al., 2013) based on MODIS NDVI and LST time series. Then different vegetation indices or indices related to water conditions were derived in order to assess the maize vegetative biomass and the effects of agricultural droughts on final grain yields. The indices are the CWSI (Crop Water Stress Index), the TCI (Temperature Condition Index), the TVDI (Temperature Vegetation Dryness Index), and the SMADI (Soil Moisture Agricultural Drought Index). Raw NDVI and SSM were also considered. Each index was then integrated over the vegetative and productive phenological phases. These phases were calculated using the dates of Start Of Season (SoS), Top of Season (ToS) and End Of Season (EoS) derived directly from the NDVI time series by thresholding. The vegetative phase corresponds to the period between the SoS and the ToS, while the productive phase corresponds to the period between the ToS and the EoS. All derived indices were re-sampled to 4-km to fit the spatial resolution of the crop model outputs.

METHODS

SARRA-O CROP MODEL SIMULATION

The SARRA-O crop model (Baron et al., 2005) was used in this study to simulate vegetative biomass at flowering, the water stress coefficient (Cstr) and attainable maize final yield over the study area, for each growing season. Each variable were simulated between 2011 and 2015 according to agricultural practices (local maize cultivar, sowing dates and fertility level), soil type and rainfall. The crop model was run using agrometeorological data from ECMWF and TAMSAT rainfall data.

STATISTICAL MODELS AND STRATEGY

To build a remote sensing based model for maize yield estimation, we adopted an ‘uncalibrated approach’ using outputs of the SARRA-O crop model. We first built a model to estimate the vegetative biomass at flowering stage based on integrated remote sensing indices over the vegetative period. Then a model to assess the crop water stress over the sensitive productive phases was built using water-related remote sensing indices. Modelled vegetative biomass and the crop water stress coefficient were then used to estimate final maize yields. Two statistical models were tested for each component of the final yield: (1) a multiple linear regression (MLR) model and a Random Forest (RF) model. To avoid overfitting in the MLR model, the number of input variables was reduced after checking for multicollinearity using the Variable Inflation Factor. Each model was evaluated based on several parameters ($cv-r^2$, $cv-RMSE$, $cv-RRMSE$, ...) using a 10-fold cross validation approach. In addition, the LMG (Lindeman, Meranda and Gold) and Mean Decrease in Mean Square Error (MSE) approaches were used to assess the contribution of each predictor variable to the final MLR and RF models respectively. Finally, a linear regression was established between SARRA-O maize yield and, biomass and water stress coefficient derived from remote sensing. The validation of the final modelled yield was done using the 2014 to 2016 ground maize yield data collected over the study area, and aggregated at the village scale.

RESULTS

The vegetative biomass at flowering stage simulated by the SARRA-O crop model exhibits a temporal variability with values ranging from 3500 kg/ha in 2015 to almost 7000 kg/ha in 2011. For the integrated water stress coefficient during the crop productive period the difference between years are not very marked, with 2015

experiencing less water stress with a Cstr around 81 (unitless), while 2011 and 2012 are characterized by a higher level of water stress (Cstr of 78). Thus, the final grain yield mainly reflects the level of vegetative biomass at flowering stage with simulated maize yields ranging from 2.5 t/ha in 2015 to more than 4 t/ha in 2011. These data were then used to calibrate a remote sensing based model over the 2011-2015 period. For the vegetative biomass at flowering stage, the final Multi Linear Regression model have a moderate but highly significant predictive power ($cv-r^2 = 0.44$). The remaining variables in the model are the CWSI, TCI and SSM, with SSM explaining 50% of the $cv-r^2$ according to LMG. The Random Forest model is significantly better than the previous one for vegetative biomass at flowering estimation with a $cv-r^2$ of 0.59 and a relative error in cross validation of 9.2%. According to the Mean Decrease in MSE, the most important variables are the TCI and NDVI. Concerning the Cstr estimation over the productive period, the MLR model exhibits a low but significant predictive power with a $cv-r^2$ of 0.26 with CWSI, TCI and SMADI remaining in the final model and TCI explaining 60% of $cv-r^2$. The RF model is also better than the MLR model with a $cv-r^2$ of 0.42 with the most important variables being TCI and CWSI. The final remote sensing-based models both have good potential for maize yield estimation with a $cv-r^2$ of 0.51 for the MLR model and $cv-r^2$ of 0.59 for the RF model. Overall, when compared to an independent dataset, the MLR model explains 57% of the observed maize yield variability while the RF model outperforms the MLR models and explains 63% of the yield variability (RMSE = 546 kg/ha) thus demonstrating the good aptitude of both models for estimating final maize grain yields.

DISCUSSION

The aim of this study was to propose and test an approach for maize yield estimation in a context where ground measurements are either unreliable or not timely available. For that purpose, we propose an 'uncalibrated approach' combining remote sensing images with outputs of a crop model (validated with ground data) as pseudo-ground data to estimate vegetative biomass at flowering stage and a crop water stress index to restrain the conversion from aboveground biomass to final grain yield. We show that whatever the statistical model considered, we can have good maize yield estimations when compared to observed data. However, we found that the RF model has a slight advantage over the MLR model in estimating maize yield, suggesting that the relationships between vegetation indices and yield are not essentially linear as previously assumed. Even if we are able to catch the main part of the observed maize yield variability, a yield overestimation is obtained. This is mainly due to the calibration of our model with outputs of a crop model that simulates attainable yields according to agrometeorological constraints but does not integrate all biotic or non-environmental factors that may lead to yield variations. Finally, given the high level of spatial heterogeneity of the Burkinabe agricultural landscape, the use of coarse resolution data is a strong limit of our study. We expect that the use of time series from multiple high spatial resolution sensors such as Sentinel, Venus or Planet represent a strong opportunity to strengthen and refine the approach proposed in this study. This would significantly improve the maize grain yield estimation over large areas and thus support the monitoring of food availability.